International Journal of Technological Exploration and Learning (IJTEL)
ISSN: 2319-2135
Volume 2 Issue 5 (October 2013)

Cognitive Development for Children's of "Hashu Advani School of Special Education" using Image Processing

Sharmila Sengupta
M.E.E.X.T.C
Department of Computer Engineering
VESIT, Mumbai, India

Komal Kripalani, Haider Ali Lakhani, Akash Punjabi
B.E. (Computer Engineering)
Department of Computer Engineering
VESIT, Mumbai, India

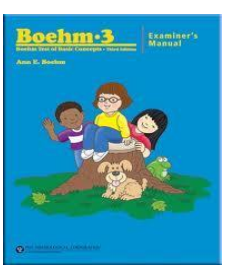
***Abstract*— Children who are deaf by birth usually find it difficult to speak because they cannot receive sound signals. To recover this ability, speech therapy of deaf children is practiced with the hope that they can speak normally. Lip reading technique is used to understand the speech of a person by visually interpreting the movements of the lips and the tongue with additional information provided by the training software. The front facial image of person is segmented and lip part is detected .This is an essential in many multimedia systems and real time applications such as videoconferencing, speech reading and understanding, face synthesis and facial animation through pronunciation .An efficient and effective detection of the lip contours from the human front facial image is relatively a difficult job in the field of computer vision and image processing due to the variation among human faces, lighting conditions and image acquisition conditions. The proposed method consists of three phases with the assumption of front facial colored image of human being as an input image. The schematic of 3Phase Model is, first phase deals with mouth localization and estimation, second phase concerns with the detection of lip contours and the last phase deals with passing the eclipse through the points of interest and then comparative study with manual lip detection is executed. Application using this technique can help the deaf people to communicate easily and motivate them further to apply it to education, research etc. An approach is proposed for the efficient development of talking and hearing skills with the help of cognitive analysis.**

Keywords- Lip Detection; Cognitive Development; Segmentation; Word Recognition.

I. INTRODUCTION

A person cannot be pushed beyond a point to learn something[7] Gaming approach has been created to develop intellectual skills among children. The book used for this purpose is as shown in fig 1 (a).

In this game, they would be able to identify the color and pronounce the corresponding color. Few snapshot of the game are as shown in which children will learn about basic concepts such as position, color, size of an object[8] etc as shown in fig. 1(b)



(a) Boehm Test of Basic Concepts

(b) Scenario of game

Figure 1. (a) Boehm Test of Basic Concepts (b) Scenario of game

More than 80% of visual information during conversation between two people is due to lip motion. Those visual clues are essential for a better understanding of the speech. But it is a challenging task because of the wide variability of lip. One of many challenges is feature extraction from image. Quality of an lip detection system depends on the feature extraction. During the last few years, many techniques have been proposed to achieve lip segmentation. Some of them use only spatial hues such as color and edges. Eveno[1] uses hue and edge information to achieve mouth localization and segmentation. There is no shape or smoothness constraint, so the segmentation is often very tough. It makes this method unsuitable for application that requires a high level of accuracy.

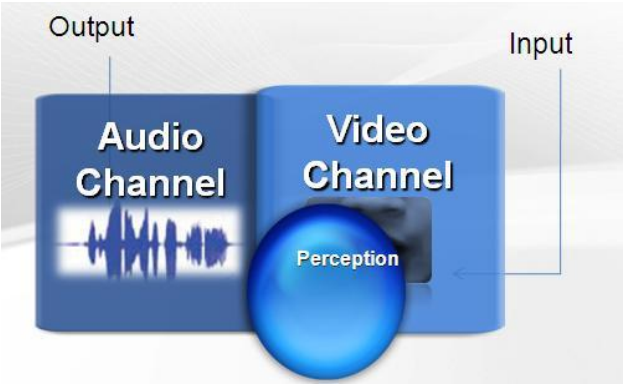


Figure 2. Lip detection approach

As shown in the fig. 2 we are using only video (images) as input and producing audio and text as output.

To make segmentation more robust and realistic, the Apriori shape knowledge has to be used. By designing a global shape model, boundary gaps are easily bridged and overall consistency is more likely to be achieved. This supplementary constraint ensures that the detected boundary belongs to possible lip shape space. For example, active shapes methods [6] can be used, but they often converge to a wrong result when the lip edges are indistinct or when lip color is close to face color. Moreover, these methods need a large training set to cover a high variability range of lip shapes.

In the fig.3 The different blocks are as follows

- Pixel based: using all pixels in lip region as feature.
- Shape based: Extract the boundary of lips as the feature
- Model based: Assume a lip modal, matching the lip shape and the modal, using some parameters to represent the shape of lip.
- Motion based: Capture the moving feature in all or parts of lip during pronunciation

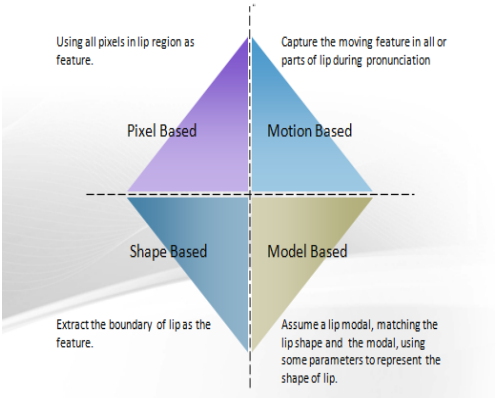


Figure 3. Word recognition approach

The proposed method is divided into three stages

(i) Face recognition [2]

(ii) Mouth region recognition [1, 5]

(iii) Key point extraction [4]

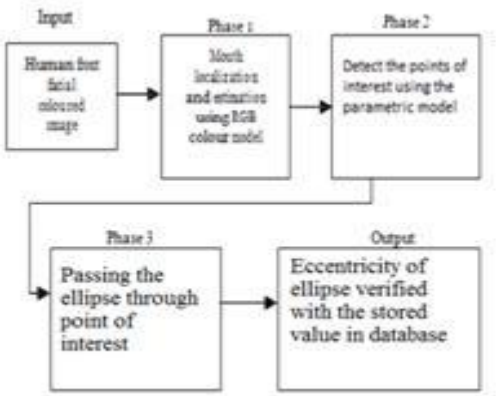


Figure 4. Schematic Diagram of Three phase Model

Description of the Three Phase Model is as shown in figure below where the first phase deals with mouth localization and estimation, second phase concerns with the detection of lip contours and the last phase deals with passing the eclipse through the points of interest and then comparative study with manual lip detection is executed.

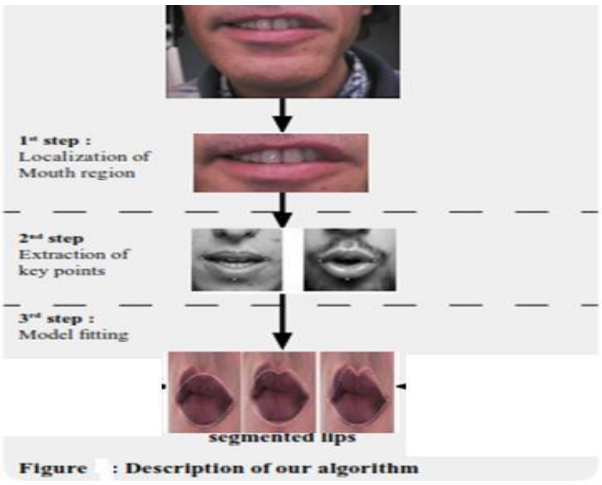


Figure 5. Desired result with high computation

The eccentricity of an ellipse is given by :

$$e=(a^2+b^2)^{\wedge 1/2}/a \quad (1)$$

Where a=semi-major axis length/width of ellipse/lip and b=semi-minor axis length/height of ellipse/lip

The array of eccentricity values will be compared with the stored values and with a certain amount of threshold; the word would either be accepted or rejected.

II. FACE DETECTION

Face detection is an initial step in different face analysis applications. Numerous face detection methods have been developed in the past years. Each method is developed in a particular context and we can cluster these numerous methods into two main approaches: image based methods and feature-based methods. The first method use classifiers trained statically with a given example set. Then the classifier is scanned through the whole image. The other method consists in detecting particular face features as eyes, nose, etc. Image of the subject taken from camera is as shown in fig. 6

(a) (b)

Figure 6. (a) Image of subject taken from camera (b) Face detection from image captured in fig. 6 (a)

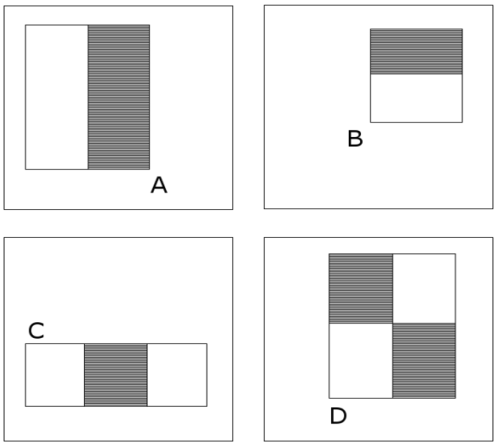


Figure 7. Basis Diagram for object Detection which is used by Viola Jones

The features employed by the detection framework universally involve the sums of image pixels within rectangular areas. As such, they bear some resemblance to Haar basis functions, which have been used previously in the realm of image-based object detection.[3] However, since the features used by Viola and Jones all rely on more than one rectangular area, they are generally more complex. The figure at right illustrates the four different types of features used in the framework. The value of any given feature is always simply the sum of the pixels within clear rectangles subtracted from the sum of the pixels within shaded rectangles. As is to be expected, rectangular features of this sort are rather primitive when compared to alternatives such as steerable filters. Although they are sensitive to vertical and horizontal features, their feedback is considerably coarser. However, with the use of an image representation called the integral image, rectangular features can be evaluated in constant time, which gives them a considerable speed advantage over their more sophisticated relatives. Because each rectangular area in a feature is always adjacent to at least one other rectangle, it follows that any two-rectangle feature can be computed in six array references, any three-rectangle feature in eight, and any four-rectangle feature in just nine.

The evaluation of the strong classifiers generated by the learning process can be done quickly, but it isn't fast enough to run in real-time. For this reason, the strong classifiers are arranged in a cascade in order of complexity, where each successive classifier is trained only on those selected samples which pass through the preceding classifiers. If at any stage in the cascade a classifier rejects the sub-window under inspection, no further processing is performed and continue on searching the next sub-window (see figure at right). The cascade therefore has the form of a degenerate tree. In the case of faces, the first classifier in the cascade – called the attentional operator – uses only two features to achieve a false negative rate of approximately 0% and a false positive rate of 40%. The effect of this single classifier is to reduce by roughly half the number of times the entire cascade is evaluated.

The face detector should only detect the face of the right person i.e the subject using the software , as the implementation is for one person the face detected should be one .

There may be multiple faces in the environment, Since the user would be the closest to the video recording device, The apparent size of his face is the largest hence in our implementation the face with the maximum area is the one which is considered.

A. Mouth Region Localization

1) Skin and Lips Color Analysis

In RGB space, skin and lip pixels have quite different components. For both, red is prevalent Moreover there is more green component than blue in the skin color mixture and for lips these two components are almost the same . Skin appears more yellow than lips because the difference between red and green is greater for lips than for skin. Hulbert and Poggio propose a pseudo hue definition that exhibits this difference. It is computed as follows:

$$H(x,y)=R(x,y)/ [R(x,y) + G(x,y)] \quad (2)$$

where R(x,y) and G(x,y) are respectively the red and the green components of the pixel (x.y). Unlike usual hue, the pseudo hue is bijective. It is higher for lips than for skin [1] Luminance is also a good clue to be taken into account, In general, light comes from above the speaker. Then the top frontier of the upper lip is very well illuminated while the upper lip itself is in the shadow. At the opposite, the bottommost lip is in the light while its lower boundary is in the shadow. To combine color and luminance information, we introduce the "hybrid edges computed as follow.

2) Ellipse application

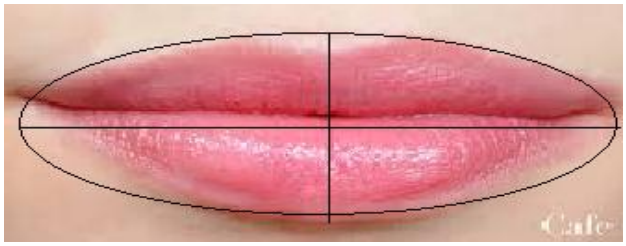


Figure 8. showing lip with an ellipse around it

The ellipse is the shape that comes closest to the shape of lip hence we use this geometric shape for our purpose, and the property of ellipse that we are going to use is the 'eccentricity'.

The eccentricity of an ellipse is given as shown in eq.1

Hence it deals with the property as the ratio of width by height of the lip.The eccentricity of a circle is zero which is nearly the shape when the speaker says 'o' and eccentricity of a lip at normal position where the width of the lip is more than thrice the height of the lip is around 0.7.

And when the speaker speaks a word which has an 'e' sound the eccentricity is nearly 0.9.

Hence when the person speaks a word like 'baby' Initially the lip is at position as above then the lip 's height increase and its width decreases hence the ellipse eccentricity decreases nearly becoming zero and afterwards the lips have increase in width hence the eccentricity increases these values can be stored in a database and then compared with a threshold.

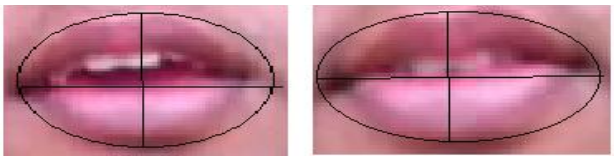


Figure 9. **S**howing two lip position when saying different syllable

Although in the figure above the difference is not that visible the result has a difference of eccentricity of 0.1 which is quite noticeable.

The above method can be reinforced by taking area of the region with outer lip as the boundary which would increase the accuracy of the algorithm.

a) Detection of the Mouth

The mouth part shown in fig 3 is obtained by using Viola Jones technique for mouth detection .The reason for obtaining the face image first was to make the lip detection more robust as the algorithm detects many things that comprise the background as mouth.

The below image consists of entire lip which is not the case with the original implementation of viola jones algorithm for mouth detection it generally cuts the upper lower lip of the person and sometimes take nose as the mouth, To avoid it we use two techniques

To avoid nose:

We consider the lower 40% of the image. Hence if M*M is the face image then the mouth region search from 6M-M in y coordinate vertically downward.

To avoid cutting of lip:

If real mouth(x1,x2,x3,x4) represents the box bounding the lip where,s

x1=x coordinate of upper left point of box.

x2=y coordinate of upper left point of box

x3=width of box. X4=height of box

Then modified coordinates would be

Real mouth(x1)=0.93*real mouth(x1); (3)

Real mouth(x2)=0.93*real mouth(x2); (4)

Real mouth(x3)=1.35*real mouth(x3); (5)

Real mouth(x4)=1.25*real mouth(x4); (6)

Figure 10. Extracting the lip region of the subject using Viola Jones for mouth region

Once the above image we subtract the pseudo-hue image from the illumination part of the image to obtain the image shown in fig.11 clearly it can be seen that the lip part is highlighted

In the image especially the horizontal corners of the lip which would help us in extracting the corner points of the lip.

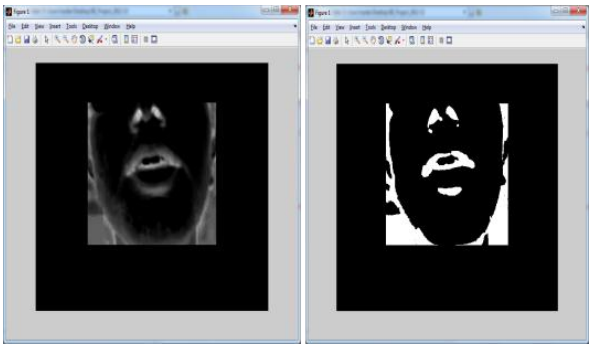


Figure 11. Obtained by subtracting the illumination component of image from pseudo-hue component and thresholding the image to obtain the black and white form of the image

An idea of how the lip is, Along with the inner contour of the lip can be obtained by converting the above image in black and white form as shown in fig.11

In the above image the region with maximum horizontal white region in the middle of the image corresponds to the upper lip. The black region in it is the opening of the mouth.

These properties would be useful in obtaining only the region of interest as shown in the fig. above.

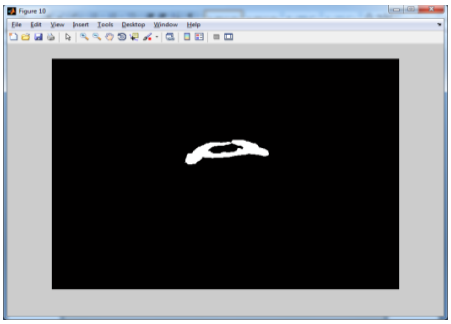


Figure 12. the upper lip of the subject with the inner opening as Black

III. Detection of Key Points

A. *The Three Upper Points*

The area of interest obtained would be sufficient to provide us with a good estimation of the lip corners along the horizontal axis.

The point which determines the top of the lip can be obtained for the black and white image would be the one having the y axis pt as the minimum of the interested region and the x coordinate as the mean of the corner pixels of the lip obtained bef ore, Due to property of symmetry in lip true for more than 70%people .The lip coordinates obtained are highlighted in the real image as shown below

Figure 13. Obtaining the corner points on upper lip

The bottom part of the lip is detected using another Region of interest in the original black and white image the topmost point in the image is the point required.

Figure 14. The lower lip region in the black and white image

Figure 15. The lower lip's points of interest

After obtaining the points an eclipse is drawn around it. And the eccentricity of the eclipse is calculated. This is done for all the images for the pronounced word and the corresponding eccentricity is saved.

B. Word tracking

Subjects were videotaped while pronouncing a word, each pronounced six times, each time as a different vowel such as A, E, I, O, U, and 'Pink', a consonant that carries an ultra-short vowel or no vowel sound as shown in figure 16. Different people pronounce the same vowels differently. Even the pronunciation of the same person in different scenarios may change.

Each word has its own sequence of mouth motion that must occur in so that the sound can be vocalized. Then the lip contour is extracted and tracked in a bounding box. The specific word associated to each tracked lip border depending on the illumination conditions.

Figure 16. sequence of image while pronouncing 'Pink'

IV. METHODLOGY

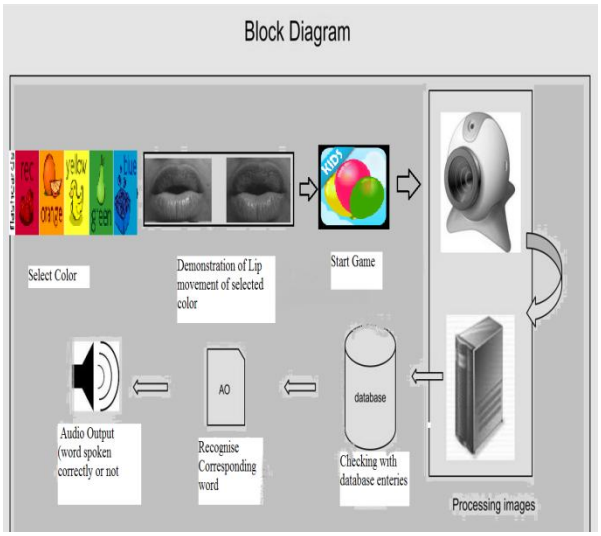


Figure 17. Block Diagram representing software

The process of the working of the software is as shown in the above fig. 17. Subject will select the color. A video demonstrating lip movement of the selected color will be played. After demonstration, game will be started where objects of different color would be there and subject need to identify the object of the color he selected and pronounce the same color. A sequence of Images of the user is acquired using web camera. The training set acquires and store word spoken by the subject. During verification, the web cam acquires real-time video for verification which is given frame by frame to an image processing unit in MATLAB. The frames are converted into binary image based on the hue of red color. The sequence of processed images is compared with a pre-existing database of analyzed images obtained. The word spoken by the subject is recognized.

The last block of the system is an additional feature for audio. If the frames pass the threshold test, the subject has pronounced the word correctly else will be demonstrated again. This concept is represented as shown in figure below:

Figure 18. Pictorial representation of game

A. Simulation Procedure

A flow chart of the steps involved in our simulation techniques is shown in fig. 19. First, the web camera acquires a fixed number of frames of the lip as the user enunciates the word. It captures a sequence of 16 or more frames using image acquisition <imaq> tool in matlab and stores it as a video file(.avi). The frames of the avi file are then extracted as jpg images per frame. The 16 by 16 frames obtained by enunciation of the word "pink" is represented in fig.16. This

is a reverse of the process that the animation industry uses to animate pictures in sequence. Each frame of this is video is now ready

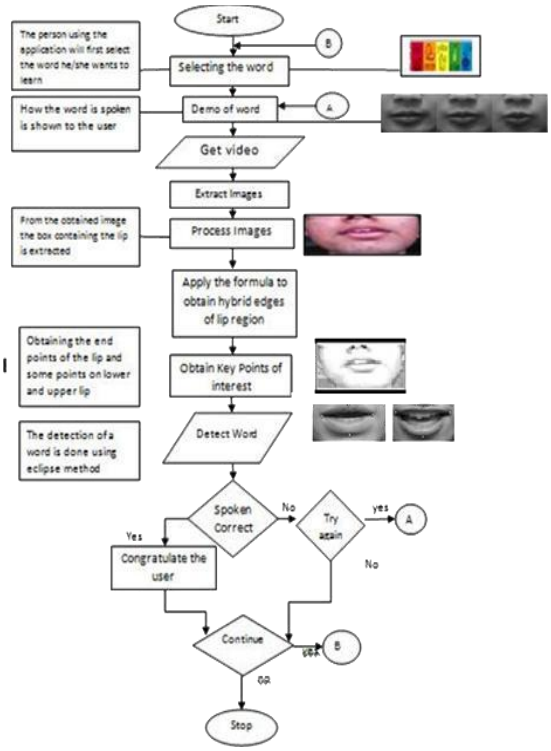


Figure 19. The Flowchart of the working of software

V. WORD DETECTION

The process follows the same routine as one carried to store the words. In addition to it, It compares the no of frames and series of eccentricities for those images to determine the word tried to be spoken by the children.

If (word said by subject == word present in the database with appropriate threshold)

%Accept the word

%go for next word till the completion of the word else

% Demonstration of the word again

VI. CONCLUSION

In this paper, an automatic, robust and accurate lip segmentation method is introduced. The use of "hybrid edges" associated with a coarse-to-fine process allow a reliable estimation of key points positions The difference in pronunciation of different people makes the lip-reading highly personalized. The software is trained based on the lip structure, complexion and features of the lip area. Environment and lighting are the limiting factors, which vary the maximum allowable difference in the threshold value. Minimizing the threshold values further enhances Moreover, Three Phase Method is introduced and implemented to detect the lip region and to extract the lip contours using Parametric Model Method. This proposed method gives the promising results for outer lip contour from the given human front facial colored images. It is an obvious that accuracy and correctness of extraction of an outer and inner, both lip contours mainly depends on clue point which is obtained through the computed hue value using red and green components of a given facial image. Making the 3PM robust along with accurate inner lip contour extraction for opened mouth is left to the future enhancements.

REFERENCES

[1] N.Eveno, A.Caplier, PY.Coulon, "'A New Colour Transformation for Ups Segmentation", iEEE Workshop on Multimedia Signal Processing (MMSP 'OJ),Cannes, France, 2001.

[2] Evangelos Skodras "An Unconstrained Method for Lip Detection in Color Images" Artificial Intelligence Group, Wire Communications Laboratory, and Nikolaos Fakotakis, Department of Electrical and Computer Engineering, University of Patras, Patras, Greece,2011

[3] Paul Viola, Michael Jones "Rapid Object Detection using a Boosted Cascade of Simple Features." IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Cambridge, 2004

[4] Ijaz Khan, Hadi Abdullah and Mohd Shamian Bin Zaina"Efficient Eyes and Mouth Detection Algorithm using Combination of Viola Jones and Skin Color Pixel Detection." University Tun Hussein Onn, Malaysia, 2006

[5] Sharmila Sengupta, Arpita Bhattacharya, Pranita Desai, Aarti Gupta Automated "Lip Reading Technique for Password Authentication",International Journal of Applied Information Systems,India,2012

[6] Ghanshyam I Prajapati Narendra M Patel,"DToLIP : Detection and Tracking of Lip Contours from Human Facial Images using Snake's Method", Image Information Processing (ICIIP), 2011 International Conference,India,2011

[7] Chuang, T.Y. You, J.H. ; Duo, A. Digital game design principles for spatial ability enhancement Dept. of Information and Learning Technology, National University of Tainan, Taiwan, Frontier Computing. Theory, Technologies and Applications, 2010 IET International Conference, Taiwan,2010

[8] Maja Rudinac, Gert Kootstra, Danica Kragic and Pieter P. Jonker" Learning and Recognition of Objects Inspired by Early Cognition" Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference , Europe,2012